Colorful brains: 14 years of display practice in functional neuroimaging

Christen, Markus ; Vitacco, Deborah A ; Huber, Lara ; Harboe, Julie ; Fabrikant, Sara I ; Brugger, Peter

Abstract: Neuroimaging results are typically graphically rendered and color-coded, which influences the process of knowledge generation within neuroscience as well as the public perception of brain research. Analyzing these issues requires empirical information on the display practice in neuroimaging. In our study we evaluated more than 9000 functional images (fMRI and PET) published between 1996 and 2009 with respect to the use of color, image structure, image production software and other factors that may determine the display practice. We demonstrate a variety of display styles despite a remarkable dominance of few image production sites and software systems, outline some tendencies of standardization, and identify shortcomings with respect to color scale explication in neuroimages. We discuss the importance of the finding for knowledge production in neuroimaging, and we make suggestions to improve the display practice in neuroimaging, especially on regimes of color coding.

DOI: https://doi.org/10.1016/j.neuroimage.2013.01.068

Title Page

Colorful brains: 14 years of display practice in functional neuroimaging

Markus Christen[a,b,], Deborah A. Vitacco[c], Lara Huber[d], Julie Harboe[e], Sara I. Fabrikant[f], Peter Brugger[c]*

[a] *Institute of Biomedical Ethics, University of Zurich, Zurich, Switzerland*

[b] *Psychology Department, University of Notre Dame, Notre Dame, USA*

[c] *Department of Neurology, University Hospital Zurich, Zurich, Switzerland*

[d] *Philosophy Department, Bergische Universität Wuppertal, Wuppertal, Germany*

[e] *School of Art & Design, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland*

[f] *Department of Geography, University of Zurich, Switzerland*

* Corresponding author: University of Zurich, Institute of Biomedical Ethics, Pestalozzistrasse 24, 8032 Zürich, Switzerland. Fax: +41 44 634 8389. E-Mail: christen@ethik.uzh.ch

Abstract

Neuroimaging results are typically graphically rendered and color-coded, which influences the process of knowledge generation within neuroscience as well as the public perception of brain research. Analyzing these issues requires empirical information on the display practice in neuroimaging. In our study we evaluated more than 9,000 functional images (fMRI and PET) published between 1996 and 2009 with respect to the use of color, image structure, image production software and other factors that may determine the display practice. We demonstrate a variety of display styles despite a remarkable dominance of few image production sites and software systems, outline some tendencies of standardization, and identify shortcomings with respect to color scale explication in neuroimages. We discuss the importance of the finding for knowledge production in neuroimaging, and we make suggestions to improve the display practice in neuroimaging, especially on regimes of color coding.

Keywords: Neuroimaging; visualization; fMRI; PET; colors; color-scale; display-style; software

Manuscript

1. Introduction

Colors are a powerful attribute of our visual world. They highlight, decorate, symbolize, regulate, soothe and warn. They mediate feelings and emotions but also influence cognition (Elliot and Maier, 2007). Perceptual illusions make us "see" non-existing colors (Benham, 1894), judge two identical colors as different or perceive different colors as identical (Wong, 2010). Rapid automatic processing of color interferes with other senses and biases the hedonic aspect of a perceptual experience (Österbauer et al., 2005), the cognitive evaluation of sentences (Reber and Schwarz, 1999), or the trustfulness of data (Skarlatidou et al., 2011). Based on such findings, visualization research has established guidelines to select appropriate colors and color scales for data display and to adjust the associated color maps for particular applications (Silva et al., 2011).

Also in neuroimaging, color has achieved a powerful role in data visualization (Ashby, 2011; Dumit, 2004; Otte and Halsband, 2006; Schott, 2010). After a century of monochrome medical imaging, its introduction has revolutionized the field. However, coloring the brain in action has been less reflected on in neuroscience than, e.g., in geographic information visualization, where major cartography textbooks discuss color perception, semantics of color, color models, and the use of color hue, saturation, and lightness to convey data relationships in maps (Kraak and Ormeling, 2010; Slocum et al., 2010). In contrast, methodological discussions in neuroimaging focus on issues like study design (Caplan, 2009), signal interpretation (Logothetis, 2008), or circular data analysis (Kriegeskorte et al., 2009). If considering the effect of images on a lay audience with respect to the persuasiveness of a scientific argument (Keehner et al., 2011; McCabe and Castel, 2008), color has not figured as a prominent attribute. The same is true for ethical concerns about misleading image-driven misrepresentations of neuroscientific results in the public (Illes et al., 2010) or courtroom (Baskin et al., 2007). Empirical research on the display practice in neuroimaging (methods of image creation, image iconographies, epistemic effects of neuroimages, etc.) is rather scarce (Alač, 2004; Beaulieu, 2002; Burri, 2008; Dumit, 2004).

Our study intends to inform the debate about neuroimaging data visualization. Based on an analysis of more than 9000 functional brain maps published over a period of 14 years we depict the display practice with respect to color use and consider underlying factors like image creation software or image production sites. Our focus was on the breadth of display styles, trends in standardization and potential shortcomings in the use of colors. This empirical grounding is important, because data

visualization is a central component in creating and advancing knowledge in scientific communities in general (Jones and Galison, 1998; Tufte, 1997) and in medicine in particular (Kevles, 1998).

As the pervasiveness of colors in today's structural and functional neuroimaging is striking (Schott, 2010), a focus on color in depicting the display practice in neuroimaging suggests itself. Colors can be used in four different ways for visually presenting complex information: to label, to represent or to imitate reality, to measure, and to enliven or decorate (Tufte, 1990). In the neuroimaging literature, examples for all four purposes are found, but colors may serve multipurpose functions and their precise role is often not clearly circumscribed (Schott, 2010). Furthermore, color is often used to picture mathematical data such as the magnitude of a statistical parameter reflecting activation differences between two or more brain regions. This raises several issues that merit consideration. First, viewers might overestimate the distinctness of research data due to display differences that are not conveyed by the data themselves. For instance, two colors might be perceived as more distinct than two different shades of the same color (Dumit, 2004). This invites the question of which display styles are used in neuroimaging. Second, the perceptual impact of colors is dependent on culture and varies with their perceived ecological valence (Palmer and Schloss, 2010). These subtle, but influential factors of color perception are largely outside an observer's awareness (Elliot and Maier, 2007). Thus, the understanding between illustrator and viewer, so crucial in the communication process, is implicit at best. Given the sheer number of color scales available (Ware, 1988), it raises the question whether a standard has emerged that would be comparable to, e.g., the field of electroencephalography (Herrmann et al., 1989). Third, even when taking for granted that color-coding does improve the way we access scientific data, the issue of standardizing perception poses major challenges. These include a raft of implicit assumptions about objectivity, the nature of the observer, the role of instruments, and the trade-offs between standardization and descriptive power (Johnston, 2002), but also involve very practical issues, such as the use of numerical scales in univariate representations of data using colors (Silva et al., 2011).

To address these issues, we investigated data of 14 years (1996-2009) of functional neuroimaging display practice in six major journals representing three different readerships (cognitive neuroimaging, imaging for clinical purposes, general scientific readership). To make the study feasible, we focused on the two most important functional imaging technologies fMRI and PET, including variants of fMRI (BOLD and non-BOLD-contrasts; Figley et al., 2010), but excluding diffusion tensor imaging and SPECT. By using a complete sampling approach, we created a database of publications and figures published in *Annals of Neurology* and *Brain* (neurology journals), *Human Brain Mapping* and *NeuroImage* (neuroimaging journals), and *Nature* and *Science* (broad interest

journals). We collected information on the origins of the contributions, on the software used in image analysis, on image structure and image complexity (Berlyne, 1971), and on the presence or absence of numerical explanations of neural activation. We coded all pictures with respect to the use of color scales in brain activation in order to identify different styles of brain images and evaluated their underlying regimes of data presentation. Coder reliability, data accuracy, and data completeness were carefully checked.

2. Material and Methods

This project extended over 33 months. Project start was in December 2010. After preparatory work, database creation and image coding extended from May 2010 to May 2011; Preliminary data analysis, completeness check, and data cleaning from June 2011 to February 2012; and data analysis and paper writing from March to September 2012. In the following, we describe the methodology in detail.

2.1 Choice of imaging methods, journals, and timeframe

In neuroimaging, many methods addressing structural (e.g., magnetic resonance imaging, MRI; diffusion tensor imaging) or functional (e.g., positron emission tomography, PET; single photon emission computed tomography; functional MRI, fMRI) aspects of neuronal systems are employed. As functional neuroimaging raises specific issues with respect to the use of colors and because we wanted to understand the temporal development of the display practice, we restricted ourselves to images created by fMRI and PET.

We expected differences in the display practice in different scientific communities (Dumit, 2004). Therefore, we chose journals that either fell into the subject category “neuroimaging” or “clinical neurology” of the *Thomson Reuters Web of Knowledge*® classification. For each category, we took the journals with the highest and second-highest impact factor (IF) based on the 2009 data that have been published at least since the mid-1990s. We selected *Human Brain Mapping* (IF = 6.256) and *NeuroImage* (IF = 5.739) for the category neuroimaging, and *Brain* (IF = 9.490) and *Annals of Neurology* (IF = 9.317) for the category clinical neurology. Lancet Neurology (IF = 18.126) has been excluded as the journal’s first issue was in 2002. We note that there is no one-to-one mapping between journals and community. For example, *Human Brain Mapping* and *NeuroImage* also rank at positions 3 and 4 for journals of the subject category “Radiology, Nuclear Medicine & Medical Imaging”, which indicates that representatives of clinical imaging also publish in these journals. Due

to these overlaps, we do not intend to map in detail all scientific fields in which functional imaging has gained importance (e.g., psychiatry; Fusar-Poli & Broome, 2006; see also Discussion).

In addition, we have chosen the journals *Nature* (IF = 34.480) and *Science* (IF = 29.747) as representatives of journals that address a broad scientific community. As all papers were saved as pdf-files, the access rights of the six journals in the library of the University of Zurich determined the beginning of our time span. Building up the database started in 2010. Thus, our time frame spanned from January 1996 to December 2009.

2.2 Data acquisition and completeness tests

Data acquisition was performed in two steps. First, a publication database was created. This database consists of contributions of all kinds (including editorial articles in *Nature* and *Science*) from the six journals that contained figures displaying data generated either by fMRI or by PET, and in which a brain or brain part picture is shown. We pursued full sampling by consecutively examining the electronic table of contents of each journal issue. If a potential hit based on the title and the abstract was identified, the pdf file was downloaded and checked for figures. If the contribution contained appropriate figures, it was included into the database. Some contributions in *Human Brain Mapping* (~5%) or *NeuroImage* (~6%) were very technical and contained figures displaying fMRI or PET measurement results with no resemblance to a brain image or displayed only simulated fMRI data. Those were excluded from the data set. Supplementary online information (and the figures therein) were also not included. For creating the publication database, additional workforce beside the three coders was involved.

The image database was set up using a template with the following items: Figure identification number (consecutive numbering); reference information (journal volume, year, and first page of the contribution); title of the contribution; type of the contribution (research article, review, short communication, comment, and editorial article in the case of *Nature* and *Science*); authors; nationality of the contribution (defined by the country in which the lab of the last author is located, as the last author usually is the supervisor of the research published); name and location of the lab of the last author (usually the name of the institute or department); text of the abstract of the contribution; text of the figure caption; method used (fMRI or PET); use of colors in the figure (yes or no); software used for image creation (see below); reference of the colors used (e.g., activation location, t-values); image coding (see next paragraph); presence of a color scale (if applicable); image structure (number of parts in the figure differentiated, e.g., by a), b), c) etc., and number of single brain pictures per part).

Each figure of a contribution generated one entry in the database. A “figure” (or image) is a single picture or a set of pictures clearly denoted as a distinct entity in the contribution by the facts that it has a figure caption and that it is cited in the text as “Figure x”. In most cases, a figure consisted of more than one picture (captured by the item “image structure”) and the pictures may also involve different color scales or even different methods. Some items of an entry may therefore include several parts (e.g. two coded color scales). In rare cases, a figure consisted of a complex arrangement of pictures (e.g. flow chart) and we tried to capture the structure of the figure as accurate as possible for the item “image structure”. The image database was set up by three coders (M.C., D.A.V., L.H.) and each of them coded a randomly assigned set of journal volume numbers in order to limit the bias that the display style of a certain journal may impose.

In the majority of the cases, either the method section or the supplementary online material contained a paragraph describing the data analysis and the software used for performing the analysis. In most cases (in particular for the common tools like SPM, AFNI, and BrainVoyager), we can assume that the software was also used for creating the single pictures, although information on image post-processing (using graphics software like, e.g., Adobe Illustrator®) that is necessary for merging different types of pictures (brain images, charts, etc.) into a single figure is not outlined in any contribution. For some software or in-house programs that have not been specified further, we were unable to clarify whether they were actually used for generation of pictures suitable for publications. As this is rare, we did not consider this shortcoming relevant.

It turned out that identifying fMRI and PET papers based on the electronic table of contents was more difficult in *Brain*, *Annals of Neurology*, *Nature* and *Science* than in *Human Brain Mapping* and *NeuroImage* due to the lower number of those papers compared to the total number of papers per issue. Furthermore, editorial contributions in *Nature* and *Science* were not always accessible as pdf file, impeding the identification of articles with brain images. Therefore, we repeated the full-text scanning by manually skim-reading all issues in the printed editions of those journals from 1996 to 2009 – also because the number of papers and images was significantly smaller than for *Human Brain Mapping* and *NeuroImage*, i.e. a failure in missing a paper would have a higher statistical weight. In this way, between 12% (Nature) and 31% (Brain) more contributions have been identified and added to the publication database.

In order to assess the completeness of our database, we calculated for each journal (excluding *Nature* and *Science*) the number of all contributions between 1996 and 2009 that included the terms

"PET" or "fMRI" (or related expressions; full-text search) and compared this number to the number of contributions in our publication database. The numbers reveal coverage of 100% for *Annals of Neurology*, 92% for *Brain*, 83% for *Human Brain Mapping* and 72% for *NeuroImage*. By taking into account that a term-based search also counts publications that do not contain fMRI or PET images (about 8%, if *Brain* is taken as a proxy) and that in *Human Brain Mapping* (5%, see above) and in *NeuroImage* (6%) contributions that were too technical have been excluded, we can conclude that we reached full coverage for *Annals of Neurology*, almost full coverage for *Brain* and that we missed between 4-14% of all fMRI and PET contributions in *Human Brain Mapping* and *NeuroImage*. As the number of images for both journals is very high and as the missed contributions are probably randomly distributed in the time interval 1996 to 2009, we can assume that completeness of the data set is sufficient for a robust trend analysis.

2.3 Image coding and coder reliability check

The color use in images has been coded as follows. The base colors have been attributed numbers: black: 1, white: 2, red: 3, yellow: 4, green: 5, blue: 6, violet/purple: 7. If an image used colors merely to pinpoint specified regions, e.g. using red and blue for this purpose, we coded this as 3/6. If an image used, e.g., a full rainbow scale to illustrate, e.g., t-values, we code this as: 7-6-5-4-3. In the sequence, we always started with the color that indicates the value closest to the statistical baseline (or lowest significance) and ended with the color that indicated the most statistically significant value. If the figure contained no explicit scale but nevertheless used color sequences to display activated regions, we zoomed into the picture and wrote down the sequence of colors beginning at the border of the colored area up to the center, e.g. an area where red is the outmost color, followed by yellow, and white is in the center, we wrote: 4-3-2. Intensity changes in colors are indicated by (1) for darker or (2) for lighter, e.g. a transition from dark green to light green is coded as 5(1)-5-5(2). Beside the colors used to label specified brain regions or to display activated areas, we also coded the color of the brain template used and the color of the background in which the template has been embedded in the picture. Usually, the template is a standard grey brain, which is indicated by the term "grey". In glass brains, the template is usually white or sometimes black with a black (or white) contour line sketching a brain. The background is usually black, sometimes white or grey.

In order to assess coder reliability, 5% of all figures in the database (460) were randomly chosen and re-coded by one coder (M.C.). Seven cases (1.52%) were identified where coding errors led to a wrong attribution of a figure to a main color scale class (see **Results**). Another 15 cases (3.26%) have been identified where not all colors in a single-color-labeling were coded (8), where a present color-

scale was not coded at all (3) or where colors have been coded using wrong numbers (4). This leads to a total error rate estimate of 4.78%, which we consider as acceptable for a robust trend analysis.

2.4 Statistical analysis

We used basic descriptive statistics and correlation analysis for the most part of data analysis using Mathematica®, Version 8.0. Inspired by the notion of image complexity by Berlyne (1971), we calculated the complexity of an image as

$$C_i = \frac{n \times l \times s}{\max(n_{l_i})}$$

where n is the total number of single pictures in the image, l is the number of levels of the image (e.g., if it consists of three parts labeled as a), b) and c), then $l = 3$), s is the number of different main color scales of the image and $\max(n_{l_i})$ is the number of pictures of the level l_i that has the highest number of pictures (taking into account that sequences of pictures on a level often show either a brain from different perspectives or a sequence of brain slices, i.e. tell "the same story" thus diminishing the complexity of the picture). As both the n and the l distributions have a pronounced long-tail characteristic, a direct comparison of the C_i-values or calculating mean values makes little sense. Instead we calculated whether a specific image is in the first two quartiles, the third or the fourth quartile of the C_i-distribution in order to analyze the trend.

3. Results

3.1 Description of the data set

Using the methodology outlined in the section **Material and Methods**, we created a publication database of 3,993 contributions and an image database of 9,179 figures (**Tab. 1**). The data confirm the expected pervasiveness of the use of colors in neuroimaging result presentations and the expected dominance of fMRI over PET.

INSERT TABLE 1

A striking feature of the data set is that few image production sites and image creation software systems dominate: Although 29 countries have been identified to host image-producing sub-units (institutes, departments and the like), only three of them (USA, UK, Germany) produced 65.6% of all images. In the temporal development, this dominance decreased over time (**Fig. 1a**). Within the top-three, the importance of Germany increased and UK sub-units produced fewer images over time.

This dominance pattern is reflected on a smaller geographical scale. 320 cities have been identified to host imaging producing sites, but London is the undisputed "imaging capital" of the world, as 11.5% of all images have been produced there. The number of identified universities and other institutions is 491; and the number of institutional sub-units is 1339. On all levels of spatial organization, a pronounced long-tailed distribution of the number of images produced per unit is discernible, i.e. few sub-units produce many images, whereas many sub-units produce only few images (**Fig. 1b**). Apart from the countries, the scaling of the distributions is comparable and the lab distribution fits a power law with exponent $k = -0.66$. This dominance pattern influences also the geographical distribution of the papers published in the journals investigated (**Fig. 1c**). In particular in the high-ranked interdisciplinary journals, papers originating from imaging sites in the USA and UK dominate (Nature: 84.1%, Science: 69.3%).

INCLUDE FIGURE 1

The same dominance pattern holds for the software that has been used for analyzing the imaging data and (supposedly, see **Discussion**) for creating the images. Although we identified 72 different software programs (excluding those labeled as "in-house software"), 85.8% of all images from which information about software was available were produced by only three systems (**Fig. 2a**, left): Statistical Parametric Mapping, SPM (originating from the *Wellcome Trust Centre for Neuroimaging*, London, UK); Analysis of Functional NeuroImages, AFNI (originating from the Medical College of Wisconsin, Milwaukee, USA); and BrainVoyager (Brain Innovation B.V., Maastricht, The Netherlands). Both SPM and AFNI are distributed freely under the Gnu General Public License Agreement, which partly explains the success of those two systems. The "quasi-monopoly" of SPM, dominating 69.3% of the "imaging market", has diminished over the years (Fig. 2a, right). Noticeable is the substantial decrease of the fraction of images where information on the analysis software was missing (49.0% in 1996, 7.2% in 2009; **Fig. 2a**, right), indicating improvements in communication standards. PET papers are more likely not to include information about the analysis software compared to fMRI paper (27.0% versus 18.6%).

INCLUDE FIGURE 2

For SPM, we (if available) collected information on the version used. Most common in our data was SPM99 (released in January 2000), followed by SPM2 (released in 2003) and SPM5 (released in December 2005; see the SPM homepage http://www.fil.ion.ucl.ac.uk/spm/; **Fig. 2b**, left). We also depicted the replacement of SPM versions (**Fig. 2b**, right). It took 2 years until SPM99 entailed about

half of all SPM-produced images and another 2 years for complete domination. The same pattern is observable for SPM2, whereas SPM5 needed much longer (4 years) until about half of all SPM-produced images relied on this version.

Finally, we also found variances within the scientific communities represented by the three journal groups. PET images are much more common in the clinical neurology journals compared to the neuroimaging journals; surprising was that the number of PET images published in *Nature* was almost twice as high as in *Science* (34.3% versus 18.3%). The number of fMRI and PET papers compared to the total number of papers per year is also substantially smaller for the clinical journals compared to the neuroimaging journals; surprising was the difference between the two clinical journals, as the percentage of those papers in *Brain* was more than double than in *Annals of Neurology* (14.4% versus 5.9%). Distinctions among *Brain* and *Annals of Neurology* were also discernible with respect to the mean number of figures per contribution (significantly smaller in Annals of Neurology) and display styles (figures in the RBS style, see below, were more common in Annals of Neurology resulting from the higher fraction of PET images).

3.2 Display styles

The color coding reveals a remarkably diverse use of colors. We identified five main display styles (**Fig. 3a**) that vary with method and software used (**Fig. 3b**). Most common is the heated body scale (HBS) in which the luminance increases from black through red, yellow and white. 44.4% of all images used this style (either the full scale or sections), the transition red-yellow (20.4%) is the single most common color scale used. The second most common styles (22.4%) are single color maps (SCM) for denoting, e.g. activated regions. Third is the rainbow scale (RBS, 15.2%), where the hue is varying in the order of the spectrum (violet-blue-green-yellow-red; or sections of this sequence), sometimes including black and white at either end. About equally frequent (6.8% and 7.4%) are single color luminance changes (CLC, e.g. from blue to white) and glass brains (GLB) – a characteristic display style for statistical maps and a peculiar format for SPM using usually black of grey shades for localizing activation and displaying the brain as a (usually) black contour on a (usually) white background. 3.9% of all images used color scales that were different from those of the five main groups (e.g. red-blue transitions).

The phenomenology with respect to the use of display styles is, however, larger than these main classes suggest, as one has to take into account that the main display style groups consist of various sub-scales. For example, only 49.2% of the scales in the rainbow scale (RBS) group used the full rainbow spectrum in the standard way (i.e. red represents the highest value of statistical

significance), 33.7% used only a section of the spectrum and 17.1% reversed the sequence of the colors (e.g., blue denoted highest significance).

The arrangement of images within a single figure is captured by our image complexity measure. The distribution is pronouncedly long-tailed, i.e. only a few images have a very high complexity (data not shown). We found no statistically significant trend in the temporal development of image complexity, i.e. the figures published in 1996 have basically the same image complexity as those in 2009, but we found expected differences for the three communities. The fraction of pictures with high image complexity (fourth quartile of the distribution) is lowest for *Science* (16.7%) and *Nature* (20.7%), higher for *Annals of Neurology* (25.2%) and *Brain* (26.0%) and highest for *NeuroImage* (27.8%) and *Human Brain Mapping* (28.9%).

INSERT FIGURE 3

3.3 Standardization

Despite the rather diverse phenomenology with respect to the use of display styles, the temporal development shows a trend of standardization towards the heated body scale (HBS) (linear regression: $r = 0.025$, $p < 0.001$), mainly at the expense of glass brains (**Fig. 3c**). This trend becomes more pronounced, if only display styles are considered that map an interval of numbers into the color space (HBS, RBS, CLC other). HBS then accounts for 63.2% of all cases.

With respect to the labeling of the color (i.e. what the color scale denotes in the image, see figure legend), also some trends of standardization are discernible (**Fig. 3d**): The “HBS standard” is most pronounced when the color scale codes for a statistical parameter explicated as such. RBS is a “quasi standard” for figures referring to “binding” (i.e. PET), and SCM is (not surprisingly) mostly used for “area” coding. Interestingly, no substantial changes in color labeling over time are discernible that could explain the overall trend towards the HBS scale, although one would expect that the largest labeling group “statistical values” (i.e., a precise labeling of a color scale) should display an increase over time, whereas the second largest group “activity” (a much more general description) should display a decrease if the communication practice would improve. However, the corresponding linear regressions are small and not statistically significant (“statistical value”: $r = 0.003$, $p = 0.40$; “activity”: $r = -0.003$, $p = 0.45$), i.e. the trend towards HBS emerged independently of changes in labeling practice.

In neurophysiological brain mapping and EEG it was agreed by convention that red and yellow indicate high activity and positive polarity, whereas green and blue are used for low activity and negative polarity (Herrmann et al., 1989, Schott, 2010). To test whether this convention also holds for neuroimaging, we counted the appearance of the basic colors red, yellow, green, blue, and violet in figures of the labeling classes (see above) “activation” and “deactivation”. A special set of figures are those that included two color-scales that coded for “activation” and “deactivation”. For this set, the convention was clearly fulfilled, i.e. activation is coded almost exclusively using red and/or yellow, and deactivation is coded using blue and/or green and/or violet. However, as soon as the figures referred to “activation” or “deactivation” alone, the standard eroded. In particular, in 54.2% of the cases “warm” colors (yellow, red) have been used for coding “deactivation” (data not shown).

A clear standardization is discernible with respect to the color of the brain template and the background color of a figure. 86.3% of the figures used a grey template and a black background, 7.5% a grey template and a white background, and 3.7% didn’t show any template (i.e. the color scale used colored the whole brain) and used a black background. Other combinations were very rare (e.g., only 1.5% of the figures used a color as background, i.e. not black or white).

Finally, also on the level of institutional sub-units, tendencies of standardizations are discernible, as laboratories/departments tend to have unique styles. To detect such trends of standardization with respect to display styles, we analyzed all sub-units that produced at least 10 images (n=259) and calculated for each of them the ratio of each display style per sub-unit. As the distributions of display style fractions – with exception of HBS – failed to pass common tests for normality (Anderson-Darling, Cramér von Mises, Pearsons χ^2), a comparison using means is not appropriate. Rather, we analyzed for each sub-unit whether the fraction of a specific style lies in the fourth quartile of the distribution, indicating a strong emphasis of a style. Whenever this was the case, the sub-unit got one point for uniqueness (for HBS, the sub-unit got also a point of the fraction lied in the first quartile, indicating a significant neglect of the generally dominating style). In this way, a measure for uniqueness is created. 15 sub-units were distinct with respect to at least 4 styles using our degree of uniqueness measure (see **Fig. 4a**). A closer look to those sub-units (**Fig. 4b**) reveals that all of them resist to use the dominating style HBS, but put an emphasis on RBS, SCM or used uncommon color scales.

INSERT FIGURE 4

3.4 Shortcomings

A striking finding is that in 38.2% of the images that displayed neuronal activations using color scales the colors were not associated with numbers by either a scale bar or by outlining the meaning of the colors in the figure caption. If the display style "glass brain" is included in this figure (when grey shadings reflect numerical data in an unspecified way), it rises to 40.9%. Missing scale explication is more common in PET than fMRI papers (46.3% versus 39.5%). Also the software used has an influence on the explication of the scale (**Fig. 5a**). In the temporal development, a tendency (although no clear trend) to explain scales is discernible: in 1996, 47.6% of the figures did not have explicated scales; the number rose up to 64.9% in 2000 and then dropped to 35.0% in 2009. This tendency seems to be influenced by software improvements, as newer versions of SPM increase the likelihood that a picture contains a scale, whereas SPM5 also increases the likelihood of using single colors (mostly red) for indicating activation (**Fig. 5b**). Furthermore, images in the CLC style surprisingly often lack explained scales (47.5%), whereas uncommon color scales are usually accompanied by scale explications (missing explications in 19.9%).

INCLUDE FIGURE 5

Another shortcoming refers to the still considerable popularity of the RBS scale, although there are well-known problems associated with it (Silva et al., 2011): First, to some users it might not present an intuitive ordering. Second, yellow is present half way through the color scale, which means that if one is interested in depicting extreme values the middle values might interfere, since yellow has a highlighting effect being perceived as brighter than the other colors. Third, the saturation steps do not equally represent differences between numbers. Yellow has the smallest number of perceived saturation steps and users find it harder to distinguish small saturation variations for yellow than, for example, for blue.

4. Discussion

We have provided an empirical analysis of the display practice in functional neuroimaging that outlines characteristics and challenges concerning the presentation of results in neuroimaging. Our first question concerned the **variability in display style**. We found that, despite a very limited number of dominating software systems and contributing institutions, a considerable variation in display style can be observed. Several factors may account for this apparent discrepancy. First, a contribution may include images for purely illustrative purposes (e.g. a PET scan) that were not produced by the software mentioned in the methods section, increasing the variability of display styles attributed to a specific software system. The current communication practice in neuroimaging

does not allow identifying those cases. Second, the specific scientific question posed may influence the display style. Testing this hypothesis would require an in-depth content analysis of the contributions in our database, which was beyond the scope of the present work. Third, although the software used for image data analysis produces the raw picture, image-post processing may increase the variability in display styles. As image post-processing is not outlined in the method sections of neuroimaging papers, it is currently not possible to quantitatively assess the importance of this factor based on published information. This would require surveys among image producers – a study that we are currently preparing. Finally, there is a fourth factor potentially responsible for the variability of display styles and for which we present empirical support. Sub-units (laboratories/departments) may develop a "unique visual language" to distinguish themselves from other sub-units. As we have shown, sub-units indeed have clear preferences with respect to the display style they use. These preferences may result from the types of questions particular sub-units address. Disentangling this possible content-driven motivation from the more superficial wish to create an own "iconography" would require an in-depth analysis of the image creation process among different sub-units.

Our second question asked for **attempts of standardization**. Across the 14 years covered by the spotlight of our analyses, some trends emerged. First and foremost, we noted an increasing popularity of the heated body scale to denote increasing activity (or increasing statistical significance) and the use of "cold" colors for a decrease in activity (or decreasing statistical significance). However, the degree of standardization still limps far behind the one reached in, for instance, cartography and geographical information visualization and a comparison with this field may be illuminating. In topographic maps, one universal convention is to render water bodies in blue, not because they are always blue (they actually rarely are!), but because people probably easily associate the color blue with water (Robinson, 1952). With the rise of more abstract, especially statistical maps since the 19th century, a more generic approach to the uses of color has been discussed for cartographic maps and statistical graphics. The selection of color shades parallel to the progression of data values – the higher the data value, the darker the color shade – has emerged as one of the few standards by the end of the 19th century (Palsky, 1999). Since then, color progressions are by convention also used in cartography to depict quantitative data sets, for example in a thematic map, when using data at the ordinal (i.e., high-medium-low incidents of crime), interval (i.e., day temperatures in degrees Celsius), or ratio level (i.e., number of inhabitants per country) of measurement. For these kinds of data sets single-hue, bi-polar hue, complementary hue, and value progressions (i.e., light-to-dark), and more recently two- and more variable color progressions and multivariate blends are commonly used in statistical maps. Many of such cartographic conventions have been tested by time and eventually found to be successful because of their commonly accepted use (Garlandini and Fabrikant, 2009).

More recently, cartographic design conventions have been empirically assessed and found to be working as predicted (Fabrikant et al., 2010; Hegarty et al., 2010), including the principles for the systematic application of color in maps (Brewer, 1994), and statistical data representations (Brewer, 1999). Neuroimaging could take these long-standing and successful mapping principles and data visualization conventions as a starting point.

Regarding our third question, the most prominent among the **shortcomings in the use of colors** is the frequent absence of any color scale explication. More than a third of the analyzed images was not accompanied by an unambiguous reference scale linking color codes to signal strength or alteration. Using Tufte's terminology (1990), colors in this case have primarily a labeling function, i.e. they pinpoint to a specific area, but do not allow representing specific activation values. For this function, use of a single color would be sufficient. Our trend analysis shows that the neuroimaging community is increasingly aware of this inappropriate inclination towards "over-colorization": use of newer software systems (like BrainVoyager) or newer versions of SMP goes along with a more frequent display of explicit color scales, and SPM5 supports the use of single colors for pinpointing on activated regions, where arguably only a labeling function was intended by the authors. This laudable trend to prevent the dissemination of "placebic" information (Trout, 2008) should be fostered as it diminishes the seductive force of color, which typically operates outside an observer's awareness (Elliot and Maier, 2007).

The importance of explaining scales is outlined by experiences made in cartography. Based on every day activities and interactions with the world, people instinctively assume, often correctly, that higher is more and that bigger is more (Lakoff and Johnson, 1980). In this sense, the rationale for the commonly accepted darker-is-more convention in cartography is the intuition that people would naturally associate darker color shades with more of whatever is being symbolized in the map (i.e., the deeper the water column in the ocean, the darker the blue shade in the map). But this convention can conflict with semantic conventions. For example, if one were to map the average days of sunshine in Santa Barbara, California, over a year, then it is not unmistakably clear whether to symbolize more days of sunshine with darker yellow (i.e., following the darker-is-more convention), or to apply a lighter yellow shade for more sunshine, following the semantic connotations of more light yielding a brighter day. Similarly, using brain images, it is not intuitively clear whether more activity should be represented by brighter or darker shades of a color. Depending on the chosen semantic category (or metaphor), say temperature, one would have to pick a fitting hue that also depends on the background color choice (i.e., light or dark). To avoid ambiguity

in the interpretation, the meaning of the color progression needs to be communicated clearly with a legend.

As outlined in the introduction, graphical, diagrammatic as much as image-based rendering of data not only is a central component in creating and advancing knowledge, but also in communicating experimental findings within a given scientific community. Research on regimes of data visualization and the specific intelligibility or cognitive accessibility of representational formats is highlighted against the abstractness of mere numerical scientific data (Delehanty, 2010; Krohn, 1991). It has even been suggested that representational formats generate a certain authority and strength of persuasiveness, which grows out of its analytical power, its power to suggest and to communicate (Giere, 1988) – a claim that has been empirically confirmed by Keener and colleagues (2011) for different types of brain images. The case has been made that the question whether representational standards or conventions bear this kind of authority has to be discussed against the background of instrumental preconditions (Lynch and Woolgar, 1990) as much as aesthetic aspects (such as symmetry or color scheme) that impact on representational formats to a significant degree, and hence might guide scientific perception and interpretation of underlying data (Huber, 2011). For instance, take a closer look at the correlation of color-code and phenomena of interest, especially red-yellow as “highness” and green-blue as “lowness” of metabolic processing. As a result, the dynamics of color perception coming along with the epistemic power of a validated unit of comparison (standard) might pave the road for an deeper understanding of the phenomena presented, hence setting the enquiry into the phenomena in question (e.g. “activity”) more and more aside.

5. Conclusions and Recommendations

In summary, our results and considerations lead to the following suggestions with respect to the display practice in neuroimaging (see also **Fig. 6**):

- The process of image production should be discussed in more detail in the method section of publications, including choice and application range of post-processing software.
- If color scales are used in images, they need to be clearly explicated by a scale or an appropriate description in the figure caption. If a figure primarily serves to display sites of activation (or the like), single colors should be preferred.
- The discerned trend of standardization with respect to using the heated body scale and “cold colors” (green-blue-transition) for increase or decrease of statistical significance should be

advanced further. However, one has to take into account that the heated body scale conflicts with the established more-is-darker principles discussed earlier, in particular if white is used to denote highest activation. Given the highlighting effect of yellow in perception, we recommend that highest statistical values should be denoted by yellow and not by white.

- Whenever possible, color scales should be decoupled from the mere labels “activation” and “deactivation”. Denoting the precise statistical meaning of the scale, or referring to a more neutral wording like “signal change” would be more appropriate.
- Non-standard displays of data relations in neuroimages (e.g., latency times etc.) should be based on color scales other than HBS and they should follow the established convention of higher equals darker color shades, and lower equals lighter shades.
- The use of the rainbow color scale may be restricted to applications where a quasi-standard has been established, e.g. for displaying binding potentials in PET imaging.
- Producers of imaging analysis tools should support appropriate use of colors both with respect to the usability of the programs as well as instruction manuals.

We believe that these recommendations are of special importance for reviewers of functional neuroimaging contributions. They have a responsibility to ensure an improvement in practice with respect to standardization and exemplification of the process of image generation in the methods part of journals that publish imaging findings. This may be of particular importance for fields in which functional neuroimaging gains increasing importance, e.g. in psychiatry (Borgwardt et al., 2012).

INCLUDE FIGURE 6

Scientific images in general and neuroimaging data in particular are communicated beyond scientific domains, hence affecting e.g., patient-physician relations or public perceptions of the explanatory power of neuroimaging devices. Here, the question whether and how the scientific community considers the ignorance of lay people with regard to scientific images is at the heart of any future debate. The interpretative authority of images using false colors seems particularly debatable in this respect. A critical evaluation of the aims and scopes of the upcoming interest in defining standards of color-coding in neuroimaging may not only impact the way scientists reflect upon their practices of data processing and representing, but also contribute to a better understanding of the rationales of color-coding in neuroimaging as such. Our vast corpus of image data collected and scrutinized here provides a firm empirical basis for the future discussion of data visualization in neuroimaging.

Competing Financial Interests

None

Acknowledgements

We thank Alvin Chesham, Annina Gläser, Cécile Haldi, Christian Ineichen, Judith Kämmer and Myriam Nösberger for their support in creating the publication database as well as the participants of the workshop "Visual Communication of Data in Neuroimaging" that took place in Zurich on June 30th and July 1st 2011 for their valuable input. This research has been supported by grant No. R-129/09 of the Cogito Foundation, Wollerau, Switzerland

Author Contributions

M.C., D.A.V., L.H., P.B., and J.H. conceived the project and discussed all of the issues. M.C., D.A.V. and L.H. performed the image coding. M.C. wrote the paper, made all the figures and guided the project. P.B. contributed to data analysis. S.I.F. contributed to the discussion of the results. All authors contributed to the editing of the paper.

Figure Legends

Figure 1: Geographical distribution of the image-producing units. a) Temporal development of the fraction of images emerging from the USA, UK, Germany, other European countries, or other countries. b) Log-log-plot of image distributions produced by countries, cities, institutions, or institutional sub-units; the straight line indicates a power law fit of the lab distribution. c) Distribution of images emerging from the USA, UK, Germany, other European countries, or other countries among journals.

Figure 2: Software used for image production. a) Top-3 software systems with respect to the fraction of images produced by these systems. Right: fraction per system (the second percent number is calculated by excluding the images from which no information is available), left: temporal development. b) Versions of SPM. Right: the fraction of images produced by different SPM versions (excluding SPM produced images from which no version information was available), left: replacement of SPM versions.

Figure 3: Display styles and standardization. a) Distribution of the five main display styles and the class “others” among the data set (left) and examples of color scales (right). b) Relationship between display style and method (fMRI or PET, left) and software used (right): The largest fractions of PET images compared to fMRI images are discernible in the GLB (42.1%) and RGB (37.5%) style. Glass brains are preferably, but not exclusively produced by using SPM (83.5%), whereas pictures in the RGB display style often lack information about the software used. c) Temporal development of the fraction of the main display styles. d) Fraction of the main display styles according to the reference of the color. We coded any labeling of the colors presented given either in the figure caption or in the figure itself (scale) into 10 classes: “activation” (explicit wording like “active”, “increase”), “activation & deactivation” (when the wording refers both to activation and deactivation), “area” (when the wording refers to a specific area of the brain), “binding” (binding potential etc.), “correlation” (when the color scale codes for the strength of a correlation between parameters), “deactivation” (for “less active”, “decrease” and the like), “parametric map” (if the wording refers to a statistical parametric map without further explications), “signal change” (for a wording that only refers to a change in the signal without further indication), special cases (e.g. color-coding of latency times), or “statistical value” (if the color scale is explicitly said to code for F, P, T, or Z values).

Figure 4: Standardization on the level of institutional sub-units. a) Distribution of uniqueness of sub-units with respect to display styles. b) Distribution of display styles among the 15 most unique laboratories/departments. The large boxes indicate the mean fraction of each display style; the grey bars indicate the ratio of images of the correspondent display style per sub-unit.

Figure 5: Shortcomings. a) The explication of color use in dependence of the software (fraction of images where the scale was explained, where it was not explained, or where only single colors were used, per software): Images generated by BrainVoyager are most likely to contain a scale, but all systems used allow the production of pictures with or without scale. b) Influence of SPM version on explication of the color scale.

Figure 6: Suggestions. A proposal to standardize the use of color scales in functional neuroimaging that is compatible with the detected trends: For increase in (e.g. BOLD) signal, the transition red-yellow is used, for decrease the transition green-blue. If the scale has no lower bound (i.e. no minimal value of statistical significance is made explicit), the color scale should start with grey (and not black), i.e. the same color type as the standard brain template. Scales for non-standard statistical data relations (e.g. correlations, latency-times) should use SCL scales and should follow the convention “higher statistical significance equals darker color shades”, and thus should be

distinguishable from the "classical" use referring to statistically significant higher or lower (BOLD) signal (an example is shown for violet). The use of the rainbow scale should be restricted to PET binding potential, they should start with black (the standard background color of brain pictures, as binding potential images do not require a brain template), and they should not include white for denoting sites with highest binding potential.

Tables

Table 1

Assumed main readership	**Journal**	**# publications (% of all publications / % research articles)**	**# figures (% PET / % color)**
Neuroimaging	Human Brain Mapping	712 (57.4 / 99.6)	1,659 (13.0 / 90.0)
	NeuroImage	2,352 (42.2 / 99.0)	5,678 (15.3 / 93.6)
Neurology	Annals of Neurology	181 (5.9 / 98.3)	301 (60.5 / 86.4)
	Brain	448 (14.4 / 97.1)	1,089 (46.1 / 76.7)
Broad interest	Nature	96 (0.3* / 71.9)	140 (34.3 / 94.3)
	Science	202 (0.7* / 60.4)	312 (18.3 / 96.8)
Total		**3,993**	**9,179 (20.2 / 90.6)**

Table 1: Overview of the dataset: In "# publications" we indicate also the fraction of the analyzed papers compared to all papers (first number in bracket) and the fraction of research articles compared to all analyzed articles; latter include also review papers or, in *Nature* and *Science*, editorial articles (second number). In "# figures" we indicate also the fraction of PET-images (first number in brackets) and the fraction of images using color (second number). The data show that 90.6% of all images used colors, 20.2% were created using PET. The temporal developments of the total number of contributions and figures shows a steady increase until 2006, where a plateau is reached (data not shown). The temporal development of the fraction of PET images per year shows a dramatic decrease over time from 78.3% in 1996 to 9.9% in 2009 (data not shown). *upper limit, as the electronic databases on which the counts are based do not contain all articles of the editorial part for *Nature* and *Science*.

References

Alač, M. 2004. Negotiating pictures of numbers. Social Epistemol. 18(2-3), 199-214.

Ashby, F.G. 2011. Statistical Analysis of fMRI Data. MIT Press, Cambridge.

Baskin, J.H., Edersheim, J.G., Price, B.H. 2007. Is a picture worth a thousand words? Neuroimaging in the courtroom. Am. J. Law Med. 33(2-3), 239-69.

Beaulieu, A. 2002. Images are not the (only) truth: Brain mapping, visual knowledge, and iconoclasm. Sci. Technol. Hum. Values. 27(1), 53-86.

Benham, C. E. 1894. The artificial spectrum top. Nature 51, 200.

Berlyne, D. E. 1971. Aesthetics and psychobiology. Appleton Century Crofts, New York.

Borgwardt, S., Radua, J., Mechelli, A., Fusar-Poli, P. 2012. Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. Behav. Brain Funct. 8, 46.

Brewer, C.A. 1999. Color Use Guidelines for Data Representation. Proceedings, Baltimore Section on Statistical Graphics. Baltimore, American Statistical Association, pp. 55-60.

Brewer, C.A. 1994. Color Use Guidelines for Mapping and Visualization. In: MacEachren, A.M., & Fraser Taylor, D.R., (eds.), Visualization in Modern Cartography. Elsevier, New York, pp. 269-278.

Burri, R.V. 2008. Doing images. Transcript Verlag, Bielefeld.

Caplan, D. 2009. Experimental design and interpretation of functional neuroimaging studies of cognitive processes. Hum. Brain Mapp. 30(1), 59-77.

Delehanty, M. 2010. Why Images? Med. Studies. 2, 161-173.

Dumit, J. 2004. Picturing Personhood. Brain scans and biomedical identity. Princeton University Press, Princeton.

Elliot, A. J., Maier, M. A. 2007. Color and psychological functioning. Curr. Dir. Psychol. Sci. 6, 250-254.

Fabrikant, S.I., Rebich-Hespanha, S., Hegarty, M. 2010. Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. Ann. Assoc. Am. Geogr. 100(1), 17-29.

Figley, C.R., Leitch, J.K., Stroman, P.W. 2010. In contrast to BOLD: signal enhancement by extravascular water protons as an alternative mechanism of endogenous fMRI signal change. Magn. Reson. Imaging. 28(8), 1234-1243.

Fusar-Poli, P., Broome, M.R. 2006. Conceptual issues in psychiatric neuroimaging. Curr. Opin. Psychiatry. 19(6), 608-12.

Gansterer N. 2011. Drawing a Hypothesis. Figures of Thought. Springer, Wien/New York.

Garlandini, S., Fabrikant, S.I. 2009. Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. In Stewart Hornsby, K., Claramunt, C., Denis, M. & Ligozat, G., (ed.), Spatial Information Theory: Proceedings of the 9th International Conference, pp. 195-211.

Giere R. 1988. Explaining Science: A Cognitive Approach. University of Chicago Press, Chicago.

Hegarty, M., Canham, M., Fabrikant, S.I. 2010. Thinking about the Weather: How Display Salience and Knowledge Affect Performance in a Graphic Inference Task. J. Exp. Psychol. Learn. 36(1), 37-53.

Herrmann, W.M., Kubicki, S., , Kugler, J., Künkel, H., , Lehmann, D., Maurer, K., Rappelsberger, P., Scheuler, W. 1989. Empfehlungen der deutschen EEG-Gesellschaft für das Mapping von EEG-Parametern (EEG- and EP-Mapping). Klin. Neurophysiol. 20(3), 125-132.

Huber, L. 2011. Norming normality. On scientific fictions and canonical visualisations. Med. Studies. 3, 41-52.

Illes, J., Moser, M.A., McCormick, J.B., Racine, E., Blakeslee, S., Caplan, A., Hayden, E.C., Ingram, J., Lohwater, T., McKnight, P., Nicholson, C., Phillips, A., Sauvé, K.D., Snell, E., Weiss, S. 2010. Neurotalk: Improving the communication of neuroscience research. Nat. Rev. Neurosci. 11(1), 61-9.

Johnston, S.F. 2002. From eye to machine: shifting authority in color measurement, in B. Saunders, J. van Brakel (eds.), Theories, Technologies, Instrumentalities of Color. Anthropological and Historiographic Perspectives. University of America, Lanhan/New York/Oxford, pp. 289-306.

Jones, C. A., Galison, P. (eds.) 1998. Picturing Science, Producing Art. Routledge, New York/London.

Keehner, M., Mayberry L., Fischer, M.H. 2011. Different clues from different views: the role of image format in public perceptions of neuroimaging results. Psychon. Bull. Rev. 18(2), 422-428.

Kevles, B.H. 1998. Naked to the Bone. Medical Imaging in the Twentieth Century. New Brunswick, Rutgers University Press.

Kraak, M.J. Ormeling, F. 2010. Cartography: visualization of spatial data. 3rd edition. Guilford Press, New York.

Kriegeskorte, N., Simmons, W.K., Bellgowan, S.F., Baker, C.I. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neurosci. 12(5), 535-540.

Krohn, R. 1991. Why Are Graphs So Central in Science? Biol. Philos. 6, 181-203.

Lakoff, G., Johnson, M. 1980. Metaphors We Live By. University of Chicago Press, Chicago.

Logothetis, N.K. 2008. What we can do and what we cannot do with fMRI. Nature 453, 869-878.

Lynch, M., Woolgar, S. (eds.). 1990. Representation in Scientific Practice. MIT-Press, Cambridge/MA.

McCabe, D.P., Castel, A.D. 2008. Seeing is believing: The effect of brain images on judgments of scientific reasoning. Cognition 107, 343-352.

Österbauer, R. A., Matthews, P.M., Jenkinson, M., Beckmann, C.F., Hansen, P.C., Calvert, G.A. 2005. Color of scents: chromatic stimuli modulate odor responses in the human brain. J. Neurophysiol. 93, 3434-3441.

Otte, A., Halsband, U. 2006. Brain imaging tools in neuroscience. J. Physiol. Paris. 99, 281-292.

Palmer, S.E., Schloss, K.B. 2010. An ecological valence theory of human color preference. PNAS, 107(19), 8877–8882.

Palsky, G. 1999. The debate on the standardization of statistical maps and diagrams (1857-1901). Elements for the history of graphical language. Cybergeo: Europ. J. Geogr. 85, online document, accessible at http://cybergeo.revues.org/148 (last access: November 13th 2012).

Reber, R., Schwarz, N. 1999. Effects of perceptual fluency on judgment of truth. Conscious. Cogn. 8, 338-342.

Robinson, A.H. 1952. The Look of Maps. An Examination of Cartographic Design. University of Wisconsin Press, Madison WI.

Schott, G.D. 2010. Colored illustrations of the brain: some conceptual and contextual issues. Neuroscientist. 16(5), 508-518.

Silva, S., Sousa Santos, B., Madeira J. 2011. Using color in visualization: A survey. Comp. Graph. 35, 320-333.

Skarlatidou, A., Wardlaw, J., Haklay, M., Cheng, T. 2011. Understanding the influence of specific Web GIS attributes in the formation of non-experts' trust perceptions. In: Ruas, A., (ed.): Advances in Cartography and GIScience. Selection from ICC 2011, Vol. 1, Lecture Notes in Cartography and Geoinformation. Springer, Berlin, pp. 219-238.

Slocum, T. A., McMaster, R. B., Kessler, F. C., Howard, H. H. 2010. Thematic Cartography & Geographic Visualization, Third Edition. Prentice Hall, Upper Saddle River/NJ.

Trout, J. D. 2008. Seduction without cause: unvovering explanatory neurophilia. Trends Cog. Sci. 12, 281-282.

Tufte, E.R. 1990. Envisioning information. Graphics Press, Cheshire.

Tufte, E.R. 1997. Visual Explanations: Images and Quantities, Evidence and Narrative. Graphics Press, Ceshire.

Ware, C. 1988. Color sequences for univariate maps: theory, experiments, and principles. IEEE Comput. Graph. Appl. 8, 41-9.

Wong, B. 2010. Color coding. Nature Methods 7(8), 573.